Discriminator optimal transport

Akinori Tanaka
(RIKEN AIP/iTHEMS, Keio Univ.)

Introduction & Summary

- Generative Adversarial Network (Goodfellow et al. 2014)

Training stage:

Require G, D, and η_1, η_2, V_1, V_2
while unconvergent :
$G \leftarrow G - \eta_1 \nabla_G V_1(G, D)$
$D \leftarrow D + \eta_2 \nabla_D V_2(G, D)$

Testing stage:

$z \sim p_Z$, images $= G(z)$

⇒If images look real, OK. If not, do you want to train it again?

- Motivations of this paper
  1. G sometimes generate unwanted shapes.
  2. D is discarded after the training.

- Summary of this paper

  1 & 2 are solved by Optimal Transport (OT) viewpoint !

Discriminator Optimal Transport (DOT)

How to do ?

- Target space DOT

$$\mathbf{argmin}_{\tilde{x}} \left\{ \|\tilde{x} - G(z_y)\|_2 - D(\tilde{x}) \right\}$$

Initialize $\tilde{x} \leftarrow G(z_y)$
for n_{trial} in range(N_{updates}) **do**
$\tilde{x} \leftarrow \tilde{x} - \epsilon \nabla_{\tilde{x}} \left\{ \|\tilde{x} - y + \delta\|_2 - \frac{1}{K_{\text{eff}}} D(\tilde{x}) \right\}$
end for
return x

$$K_{\text{eff}} = \max \left\{ \frac{|D(x_1) - D(x_2)|}{\|x_1 - x_2\|_2} \middle| x_1, x_2 \sim p_G \right\}$$

- Latent space DOT

$$\mathbf{argmin}_{z} \left\{ \|z - z_y\|_2 - D(G(z)) \right\}$$

Initialize $z \leftarrow z_y$
for n_{trial} in range(N_{updates}) **do**
$g = prj_{p_Z}\left(\nabla_z \left\{ \|z - z_y + \delta\|_2 - \frac{1}{k_{\text{eff}}} D(G(z)) \right\} \right)$
$z \leftarrow z - \epsilon g$
end for
return $G(z)$

$$k_{\text{eff}} = \max \left\{ \frac{|D(G(z_1)) - D(G(z_2))|}{\|z_1 - z_2\|_2} \middle| z_1, z_2 \sim p_Z \right\}$$

Experiment on Target space DOT

- 2d swiss-roll

Target space DOT does work !

$$\mathbf{argmin}_{\tilde{x}} \left\{ -D(\tilde{x}) \right\}$$

EMD	0.021(05)	**0.020(06)**	0.160(22)

But once we apply it to higher dimensional data…

- CIFAR-10

Target space DOT Latent space DOT

Target space DOT does not work.
But Latent space DOT works !

Experiment on Latent space DOT

		• CIFAR-10 bare	• CIFAR-10 DOT	• STL-10 bare	• STL-10 DOT
DCGAN	WGAN-GP	6.53(08), 27.84	7.45(05), 24.14	8.69(07), 49.94	9.31(07), 44.45
	SNGAN(ns)	7.45(09), 20.74	7.97(14), **15.78**	8.67(01), 41.18	9.45(13), **34.84**
	SNGAN(hi)	7.45(08), 20.47	8.02(16), 17.12	8.83(12), 40.10	9.35(12), 34.85
	SAGAN(ns)	7.75(07), 25.37	**8.50(01)**, 20.57	8.68(01), 48.23	10.04(14), 41.19
	SAGAN(hi)	7.52(06), 25.78	8.38(05), 21.21	9.29(13), 45.79	**10.30(21)**, 40.51
Resnet	SAGAN(ns)	7.74(09), 22.13	8.49(13), 20.22	9.33(08), 41.91	**10.03(14), 39.48**
	SAGAN(hi)	7.85(11), 21.53	**8.50(12), 19.71**		

- ImageNet

	# updates=0	# updates=16	# updates=32
trial1($k_{\text{eff}}(y) = 1$)	36.40(91), 43.34	37.25(84), 42.70	**37.61(88), 42.35**
trial2($k_{\text{eff}}(y) = 1$)	36.68(59), 43.60	36.97(63), 42.85	**37.02(73), 42.74**
trial3	36.64(63), 43.55	**37.51(01), 42.43**	36.88(79), 42.52
trial4	36.23(98), 43.63	37.29(86), 42.67	**37.29(07), 42.40**

In all cases, (IS, FID)s are improved* !

*We need to tune parameters $\epsilon, N_{\text{update}}$, & to choose appropreate optimizer.

How guaranteed ?

- Step1

(*1)

$$\uparrow V_2(G, D) \leq \mathbb{E}_{x \sim p_{\text{data}}}[D(x)] - \mathbb{E}_{y \sim p_G}[D(y)]$$

increased by GAN training

$$\leq \max_{\|D\|_{Lip} \leq 1} \left\{ \mathbb{E}_{x \sim p_{\text{data}}}[D(x)] - \mathbb{E}_{y \sim p_G}[D(y)] \right\}$$

Lower bound maximized ⇒ $D_{\text{trained}} \approx D^*$

- Step2

Kantorovich-Rubinstein duality

$$\left\{ \mathbb{E}_{x \sim p_{\text{data}}}[D^*(x)] - \mathbb{E}_{y \sim p_G}[D^*(y)] \right\} = \left\{ \mathbb{E}_{(x,y) \sim \pi^*}[\|x - y\|_2] \right\}$$

π^* : Optimal Transport plan

(*2)

x **for given** $y \sim p_G$ **with** $\pi^*(x, y) \neq 0$

is determined by $\mathbf{argmin}_{\tilde{x}} \left\{ \|\tilde{x} - y\|_2 - D^*(\tilde{x}) \right\}$

(*1) For example, the usual GAN objective can be bound by

$$V(G, D) = -\mathbb{E}_{x \sim p_{\text{data}}}\left[\log(1 + e^{-D(x)})\right] - \mathbb{E}_{y \sim p_G}\left[\log(1 + e^{+D(y)})\right]$$

$\mathbb{E}_{x \sim p_{\text{data}}}[\log \sigma(D(x))]$ ↓ **for** $X \geq 0, -\log(1 + X) \leq -\log X$ $\mathbb{E}_{x \sim p_G}[\log\{1 - \sigma(D(x))\}]$

$$\leq -\mathbb{E}_{x \sim p_{\text{data}}}\left[\log e^{-D(x)}\right] - \mathbb{E}_{y \sim p_G}\left[\log e^{+D(y)}\right]$$

(*2) Because of the marginality conditions on OT plan π,

$$\mathbb{E}_{x \sim p_{\text{data}}}[D^*(x)] = \mathbb{E}_{(x,y) \sim \pi^*}[D^*(x)], \quad \mathbb{E}_{x \sim p_G}[D^*(y)] = \mathbb{E}_{(x,y) \sim \pi^*}[D^*(y)]$$

And it means $\int dy\, \pi^*(x, y) = p_{\text{data}}(x)$ $\int dx\, \pi^*(x, y) = p_G(y)$

$$0 = \mathbb{E}_{(x,y) \sim \pi^*}[D^*(x) - D^*(y) - \|x - y\|_2]$$

$$= \int dxdy\, \underbrace{\pi^*(x, y)}_{\geq 0} \underbrace{[D^*(x) - D^*(y) - \|x - y\|_2]}_{\leq 0\ (\Leftrightarrow \|D\|_{Lip} \leq 1)}$$

To make (RHS)=0, integrand should be 0 almost everywhere, i.e.

$$\pi^*(x, y) \neq 0 \Rightarrow -D^*(y) = \|x - y\|_2 - D^*(x)$$

minimum

$\|D\|_{Lip} \leq 1$

$$\leq \|\tilde{x} - y\|_2 - D^*(\tilde{x})$$